

Subgroup C avian metapneumovirus (MPV) and the recently isolated human MPV exhibit a common organization but have extensive sequence divergence in their putative SH and G genes

D. Toquin,¹ C. de Boisseson,² V. Beven,² D. A. Senne³
and N. Etteradossi¹

Correspondence

N. Etteradossi

n.etteradossi@

ploufragan.afssa.fr

^{1,2}French Agency for Food Safety (AFSSA), Avian and Rabbit Virology Immunology and Parasitology Unit (VIPAC)¹, Virus Genetics and Biosecurity Unit (UGVB)², BP53, 22440 Ploufragan, France

³United States Department of Agriculture (USDA), National Veterinary Services Laboratories (NVSL), PO Box 844, Ames, IA 50010, USA

The genes encoding the putative small hydrophobic (SH), attachment (G) and polymerase (L) proteins of the Colorado isolate of subgroup C avian pneumovirus (APV) were entirely or partially sequenced. They all included metapneumovirus (MPV)-like gene start and gene end sequences. The deduced Colorado SH protein shared 26·9 and 21·7 % aa identity with its counterpart in human MPV (hMPV) and APV subgroup A, respectively, but its only significant aa similarities were to hMPV. Conserved features included a common hydrophobicity profile with an unique transmembrane domain and the conservation of most extracellular cysteine residues. The Colorado putative G gene encoded several ORFs, the longer of which encoded a 252 aa long type II glycoprotein with aa similarities to hMPV G only (20·6 % overall aa identity with seven conserved N-terminal residues). The putative Colorado G protein shared, at best, 21·0 % aa identity with its counterparts in the other APV subgroups and did not contain the extracellular cysteine residues and short aa stretch highly conserved in other APVs. The N-terminal end of the Colorado L protein exhibited 73·6 and 54·9 % aa identity with hMPV and APV subgroup A, respectively, with four aa blocks highly conserved among *Pneumovirinae*. Phylogenetic analysis performed on the nt sequences confirmed that the L sequences from MPVs were genetically related, whereas analysis of the G sequences revealed that among MPVs, only APV subgroups A, B and D clustered together, independently of both the Colorado isolate and hMPV, which shared weak genetic relatedness at the G gene level.

Received 16 December 2002

Accepted 18 March 2003

INTRODUCTION

Avian metapneumoviruses (known as avian pneumoviruses or APVs) cause respiratory diseases and/or egg drops in species such as turkey, chicken, Muscovy or Pekin duck (reviewed by Cook, 2000). First reported in the late 1970s in South Africa (Buys & Du Preez, 1980), and subsequently in France and the UK (Giraud *et al.*, 1986; McDougall & Cook, 1986), APVs have now been described worldwide, although Australia is still considered to be free from infection (Bell & Alexander, 1990). APV has been proposed as the type species for the newly defined genus *Metapneumovirus* (Pringle, 1998), which was created within the subfamily *Pneumovirinae* to account for the APV genome containing eight genes

arranged in the order 3'-N-P-M-F-M2-SH-G-L-5', instead of 10 genes arranged in the order 3'-NS1-NS2-N-P-M-SH-G-F-M2-L-5', as found in pneumoviruses from mammals, such as respiratory syncytial virus (RSV) (Yu *et al.*, 1992).

APV subgroups A and B, designated by reference to human RSV (HRSV) subgroups, were defined originally on the basis of nt sequence divergence in the gene encoding the putative attachment glycoprotein G (Juhász & Easton, 1994). This grouping was consistent with antigenic differences recognized in ELISA, virus neutralization or with monoclonal antibodies (Toquin *et al.*, 1992; Cook *et al.*, 1993; Collins *et al.*, 1993; Etteradossi *et al.*, 1995; Bâyon-Auboyer *et al.*, 1999). Recent findings have unravelled more extensive variations among APVs. The first APV isolates obtained in the United States from turkeys with rhinotracheitis (Senne *et al.*, 1997) were shown to be genetically (Seal, 1998) and antigenically (Cook *et al.*, 1999; Toquin *et al.*, 2000) different

The sequences reported in this paper have been deposited in EMBL under the accession nos AJ457967 (SH and G) and AJ496565 (L).

from subgroups A and B. The first of these isolates, known as the Colorado APV isolate, has been proposed as the prototype of a new subgroup, C, or even of a new serotype (Seal, 2000). The first APVs isolated in ducks in France were related to this subgroup (Toquin *et al.*, 1999a). Finally, French isolates obtained in the mid 1980s, which did not fit antigenically and genetically into any of the three other subgroups, were proposed as subgroup D (Bäyon-Auboyer *et al.*, 1999, 2000). Antigenic and genetic data suggest consistently that subgroup C APV is more divergent. First, in ELISA and virus neutralization assays, some cross-reactivity occurs between all subgroups except C (Cook *et al.*, 1999; Toquin *et al.*, 2000). Second, in experimental cross-protection studies performed in specific-pathogen-free chickens or turkeys, attenuated vaccines developed from subgroups A or B APV clinically control infection of viruses belonging to the four APV subgroups (Cook *et al.*, 1995; Toquin *et al.*, 1996, 1999b), whereas prior immunization with a subgroup C virus does not prevent infection by subgroup A or B viruses (Cook *et al.*, 1999). Finally, several studies targeted at the N, P, M, F and M2 genes have shown that APV subgroup C is genetically the most divergent APV (Shin *et al.*, 2002).

Until recently, metapneumoviruses (MPVs) were only demonstrated in avian species. In 2001, a human MPV (hMPV) was identified in infants presenting with respiratory signs evocative of RSV infections, first in the Netherlands (van den Hoogen *et al.*, 2001) and then in Australia (Nissen *et al.*, 2002), Canada (Peret *et al.*, 2002) and France (Freymuth *et al.*, 2002). The order of hMPV genes is typical of MPV and the high aa identity of the proteins encoded by the N, P, M, F and M2 genes of hMPV with their counterparts in subgroup C APV (68–88% aa identity) showed that these viruses were closely related (van den Hoogen *et al.*, 2002). The hMPV genome has now been sequenced fully (van den Hoogen *et al.*, 2002) and sequence divergence among hMPV isolates suggests that several genetic hMPV lineages co-circulate in humans (Peret *et al.*, 2002). However, the relationships between the human and avian viruses have not been clarified yet, as the genes encoding two proteins that may be important for antigenicity and host range, the small hydrophobic protein (SH) and the putative attachment glycoprotein (G), respectively, have not been sequenced in subgroup C APVs.

Here we report the first sequences of the SH and G genes of a subgroup C APV, the Colorado isolate, together with the partial sequence of its polymerase (L) gene. These data are compared with their counterparts in hMPV and other APVs; although hMPV and the Colorado isolate of subgroup C APV share some features of their putative SH and G proteins, which make them clearly different from the other APV subgroups, the two viruses also exhibit extensive sequence divergence.

METHODS

Virus. The Colorado isolate of subgroup C APV was kindly provided by D. A. Senne (USDA, APHIS, Ames, USA), as reference

193ADV9802 passaged 12 times in chicken embryo fibroblasts (CEFs). It was passaged twice in Vero cells in the AFSSA laboratory prior to genetic study. The antigenicity of the Colorado isolate has been characterized previously (Toquin *et al.*, 2000).

RNA extraction, RT-PCR and oligonucleotide sequencing.

Virus RNA was extracted from inoculated Vero cells using the RNeasy Mini kit (Qiagen). RNA extracts were reverse-transcribed at 42 °C for 50 min using Superscript II RT (Invitrogen), according to the manufacturer's recommendations. All PCRs were performed with the proofreading polymerase from the Expand High Fidelity PCR system (Roche), according to the manufacturer's recommendations. The PCR cycle included a pre-cycle step at 94 °C for 2 min, followed by 15 cycles of denaturation at 94 °C for 15 s, annealing at the optimal temperature, according to the selected primers, for 30 s and extension at 72 °C for 2 min. For the next 25 cycles, the extension step was increased by 5 s per cycle. A final extension step was performed at 72 °C for 7 min. Oligonucleotide primers used for cDNA synthesis and PCR were selected in sequences published already or determined previously using the OLIGO 4.1 Primer Analysis software (Med. Prob.). Information on the sequence and optimal annealing temperature of the primers is available upon request. PCR products were purified with the GeneClean II kit (Bio 101) and were sequenced in both directions with an ABI 373 XL DNA sequencer, the AmpliTaq DNA polymerase FS, the ABI Prism Dye Terminator Cycle Sequencing kit (all from Applied Biosystems) and the PCR primers.

Strategy for sequencing the SH, G and L genes. Primers selected previously in flanking genes were used first (F+1087 and L-1207 primers; Bäyon-Auboyer *et al.*, 2000). Due to mispriming, the resulting PCR only allowed sequencing of 650 bp of the Colorado L gene. A Colorado-specific primer pair (F+354/L-765) was then selected, which amplified a 4800 bp PCR product encompassing the Colorado genomic RNA from the F to the L genes. The sequence of this PCR product was determined by genome walking and was confirmed by sequencing at least three different PCR products corresponding either to the original long PCR product or to the shorter PCR product designed during the genome walking process.

Submission of sequences. Sequences presented here are from cDNA and have been deposited in the EMBL database with accession numbers AJ457967 (SH and G) and AJ496565 (L).

Analysis of nt and aa sequences. Sequences were aligned using CLUSTAL W (Thompson *et al.*, 1994). Hydrophobicity profiles were determined with the TopPredII software (Claros & von Heijne, 1994) available online at the Institut Pasteur, France (<http://bioweb.pasteur.fr>). Proteins were analysed on the same website using the Pepstats software (EMBOSS). O-Glycosylation, N-glycosylation and signal peptidase sites were predicted using NetOGlyc2.0 (Hansen *et al.*, 1998), NetNGlyc1.0 (Gupta *et al.*, 2002) and SignalP1.1 (Nielsen *et al.*, 1997), respectively; all software derived from the Center for Biological Sequence Analysis, Technical University of Denmark, Denmark (<http://genome.cbs.dtu.dk>). Local similarities between nt and aa sequences were studied with BLAST 2 (Altschul *et al.*, 1997).

Phylogeny. Multiple alignment of APV, hMPV and pneumovirus aa sequences was performed as described above. To avoid insertion of gaps within codons, multiple alignment of nt sequences was deduced from alignment of aa sequences. Sites including only gaps introduced by the outgroup (SV5, simian parainfluenza virus type 5) were eliminated. The final alignment was analysed using PHYLIP, version 3.52c (Felsenstein, 1993). The SEQBOOT program was run to generate 500 datasets that are randomly resampled versions of the nt sequences aligned previously. Each dataset was analysed using the

neighbour-joining method with Kimura's modified two-parameter distances (DNADIST and NEIGHBOR). The resulting 500 phylogenetic trees were used to compute a consensus tree according to CONSENSE with the 'majority rule' criteria. Bootstrap values on the consensus tree are percentages based on the 500 bootstrap iterations.

RESULTS

The putative SH gene and its deduced product

The putative Colorado SH gene started immediately downstream of the published gene end sequence of the M2 gene (AGTTAATAAAAAATT; accession no. AF176592). Its length was 623 nt from the first nt of its transcription start signal to the last nt of its gene end signal. The transcription start signal was similar to APV subgroup A SH gene start (GGGACAAGT in strain CVL/14.1, which is the only APV SH gene sequenced to date; accession no. S40185) and different from that of the hMPV SH gene (GGGATAAAT in strain 00-1; accession no. AF371337). As in hMPV, the first ATG codon of the SH gene was located 4 nt downstream of the transcription start signal. A large ORF started with this ATG (525 nt encoding a 175 aa product in the Colorado isolate versus 549 nt encoding a 183 aa product in hMPV and 522 nt encoding a 174 aa product in APV subgroup A). The stop codon of the SH ORF of the Colorado isolate was located 69 nt upstream of the transcription stop signal (AGTTATTTAAAAA), two other potential stop codons being included in these 69 nt. The best nt identities with the SH sequence of the Colorado isolate were 55.3 and 52.5 % for hMPV or APV subgroup A viruses, respectively.

The deduced SH protein of the Colorado isolate had a predicted molecular mass of 19.5 kDa (versus 20.9 kDa for hMPV and 18.8 kDa for CVL/14.1) and an isoelectric point of 8.65 (8.96 for the hMPV and 7.99 for APV subgroup A). Its only significant similarities, as detected with BLAST, were with the hMPV SH protein (accession no. AF371337).

The deduced Colorado SH protein had a high serine and threonine content (22.28 %), similar to hMPV SH (22.40 %) and higher than APV subgroup A SH (17.81 %) (Fig. 1). No conserved O-glycosylation sites were observed between the Colorado and hMPV SH proteins. The extracellular domain of the Colorado SH included four N-glycosylation sites versus two (with one in a conserved position) in hMPV SH. The predicted hydrophobicity profile of the Colorado SH protein included an N-terminal hydrophilic region with a potential intracellular position (aa 1–28) followed by a short hydrophobic area, predicted to be a membrane-spanning domain (aa 29–49), and a hydrophilic C-terminal end, predicted to be an extracellular domain (aa 50–175) (Fig. 1). This hydrophobicity profile was similar to the hMPV SH protein (three predicted intracellular, transmembrane and extracellular regions, spanning aa 1–30, 31–51 and 52–183, respectively) but different from the APV subgroup A SH protein, which had two predicted hydrophobic transmembrane regions (aa 32–52 and 69–89). Although the deduced Colorado SH protein shared only 26.9 % overall aa identity with the hMPV putative SH protein (and 21.7 % with APV subgroup A), some features were conserved with hMPV SH (Fig. 1): (i) the conservation of eight cysteine residues of nine located

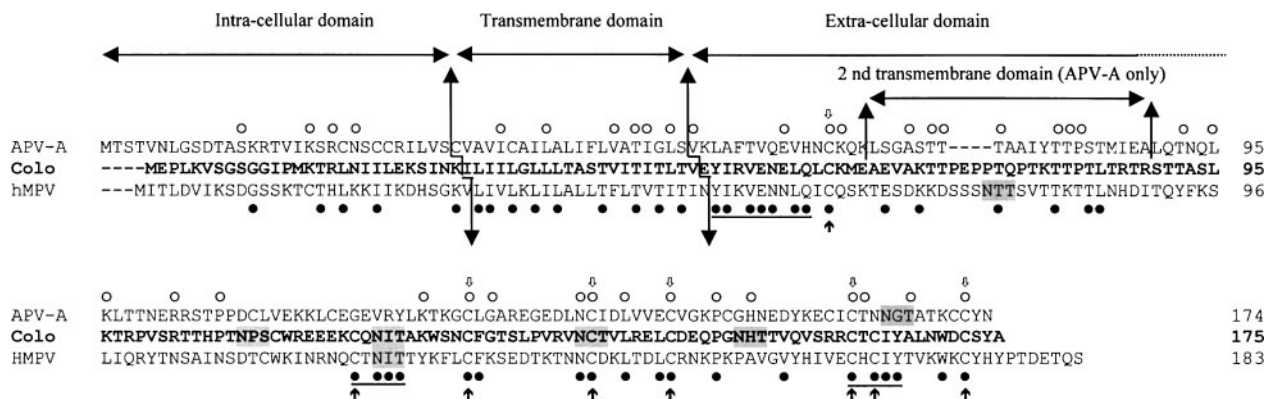


Fig. 1. Alignment of the predicted aa sequence of the SH protein of APV subgroup C Colorado isolate (Colo) with APV subgroup A strain CVL/14.1 and hMPV strain 00-1 (hMPV) (see accession numbers in the legend of Fig. 4). The alignment was done with CLUSTAL W and was presented so as to give the optimum alignment of the APV subgroup A and Colorado sequences on the one hand and of the Colorado and hMPV sequences on the other hand. Proposed intracellular, transmembrane and extracellular domains are indicated above the sequences. Open marks are used for comparison between APV subgroup A and Colorado and closed marks are used for comparison between Colorado and hMPV. Circles indicate the positions of conserved aa residues. Arrows indicate conserved cysteine residues. Predicted sites of N-glycosylation are enclosed in shaded boxes. The three aa stretches longer than four residues and with a conservation greater than 75 % are underlined. Gaps (periods) introduced into the sequence to optimize alignments are not taken into account in the arbitrary numbering.

in the extracellular domain (six of these were also conserved in APV subgroup A); (ii) the presence in the Colorado isolate and hMPV SH proteins, as in all *Pneumovirinae*, of basic aa flanking the transmembrane domain (lysine and arginine residues at aa 28 and 53, respectively, in the Colorado isolate; Fig. 1); and (iii) the presence of three short aa stretches with an identity greater than 75% [aa 51–59, 118–122 (which includes the conserved *N*-glycosylation site) and aa 162–166]. Pairwise alignment of the Colorado putative SH protein with the SH proteins of members of the genus *Pneumovirus* revealed 6.3–16.6% identity [with ovine RSV (ORSV) and pneumonia virus of mice (PVM), respectively] and no conserved aa stretch longer than 3 aa was apparent in these alignments.

The putative G gene and its deduced products

The Colorado putative G gene started 91 nt downstream of the SH gene. Its length was 783 nt from the first nt of the transcription start signal [GGGACAAGT, as found in most APV genes sequenced to date (Bäyon-Auboyer *et al.*, 2000) and in the G gene of hMPV (van den Hoogen *et al.*, 2002)] to the last nt of the gene end/polyadenylation signal. As observed also in hMPV, the Colorado G gene had the capacity to encode several polypeptides. The start codon of the larger ORF (ORF1) was located, as in hMPV, 4 nt downstream of the gene start. ORF1 was 759 nt long, hence 48 nt longer than the hMPV G main ORF, but 417, 486 and 411 nt shorter than the G genes of APV subgroups A (strain CVL/14.1; accession no. S40185), B (strain 2119; accession no. L34031) and D (strain Fr/85/1; accession no. AJ400731), respectively. In the Colorado isolate (and unlike hMPV), no shorter secondary ORF was found in the same reading frame after the stop codon of ORF1, which was partially overlapping with the gene end/polyadenylation signal (AGTTAATTAAAAA, the first 2 nt belonging to the TAG stop codon).

As in hMPV, the second reading frame of the Colorado putative G gene potentially encoded two additional polypeptides [71 and 75 aa long and encoded by nt 296–511 (ORF2a) and 521–748 (ORF2b), respectively] and the third reading frame encoded one [137 aa long and encoded by nt 330–743 (ORF3)]. However, these secondary polypeptides were neither located downstream of an identifiable gene start signal nor started with an ATG codon, and only the ORF3-encoded polypeptide ended with the typical APV AGTTA gene end signal partially overlapping its TAG stop codon.

The ORF1-encoded protein was 252 aa long (predicted molecular mass of 27.4 kDa, isoelectric point of 10.4). Its only significant similarities, as revealed by BLAST, were with the putative hMPV G protein (accession no. Q8QN55). The alignment of the aa sequence of the two proteins is shown in Fig. 2. The overall aa conservation between the ORF1-derived and hMPV G proteins was low (20.6%, corresponding to a 55.6% nt identity), but much higher in the intracellular domain (53.3% aa identity, with 100%

aa identity of the seven N-terminal residues) than in the transmembrane and extracellular domains (19.0 and 15.9%, respectively). The ORF1 protein contained 22.6% threonine, 6.7% serine and 6.3% proline residues (18.2, 15.7 and 8.5% in hMPV, respectively). Only two cysteine residues were present in the ORF1 protein and only the one located in the intracellular domain was conserved with hMPV G. A total of 16 potential O-glycosylation sites were conserved between the two proteins. The extracellular domain of protein ORF1 included five potential *N*-glycosylation sites (versus four in the hMPV G extracellular domain) but none was conserved with hMPV. The hydrophobicity profile of the ORF1 protein included an N-terminal hydrophilic region (aa 1–30) predicted to be intracytoplasmic, a hydrophobic segment predicted to have a transmembrane location (aa 31–51) and a C-terminal hydrophilic part proposed as an extracellular domain (aa 52–252) (Fig. 2). This organization is consistent with that of a type II glycoprotein and similar to the predicted hMPV G protein. Specific features of the Colorado ORF1-encoded protein included the presence in the transmembrane domain of two methionine residues (positions 45 and 48) together with a predicted cleavage site for an eukaryotic signal peptidase (between aa positions 46 and 47), and a longer C-terminal region (200 aa in the Colorado isolate versus 184 aa in hMPV). Comparison of the ORF1 protein with the G proteins of APV strains belonging to subgroups A, B and D (Fig. 2) revealed low aa identities (21.0, 19.0 and 18.3%, respectively), a lack of conserved cysteine residues in the extracellular domain and the absence of the short hydrophobic aa stretch shown to be conserved in the extracellular domain of the other APV subgroups (Bäyon-Auboyer *et al.*, 2000). Optimum pairwise alignments of the Colorado ORF1 protein with the G proteins of viruses belonging to the genus *Pneumovirus* revealed 13.1–18.3% identity with PVM and HRSV subgroup B, respectively.

Regarding the polypeptides encoded possibly by ORFs 2a, 2b and 3, the similarity searches revealed only matches of low significance. Positive matches with RNA viruses other than hMPV were found only for the ORF2a polypeptide, firstly with the N-terminal end of the virus coat protein of several Sindbis virus isolates (e.g. accession no. P03316, E value = 0.01) and secondly with the proline-rich C-terminal half of the G protein of some subgroup B RSV isolates (e.g. accession no. Q9DLA6, E value = 0.90). The BLAST search for local similarities among the polypeptides encoded by the secondary G ORFs of all MPVs (hMPV and APV subgroups A–D) revealed only matches of low significance between the Colorado virus and hMPV (the most significant match was between the polypeptides encoded by ORF2b in both viruses, E value = 0.001). Some more significant matches were detected also: (i) between the Colorado ORF2a and a polypeptide corresponding to ORF2b in APV subgroup B (E value = 2×10^{-4}); (ii) between the polypeptides corresponding to ORF2b in both the Colorado virus and APV subgroup D (E value = 2×10^{-4}); and (iii) between the Colorado ORF3 polypeptide and polypeptides corresponding

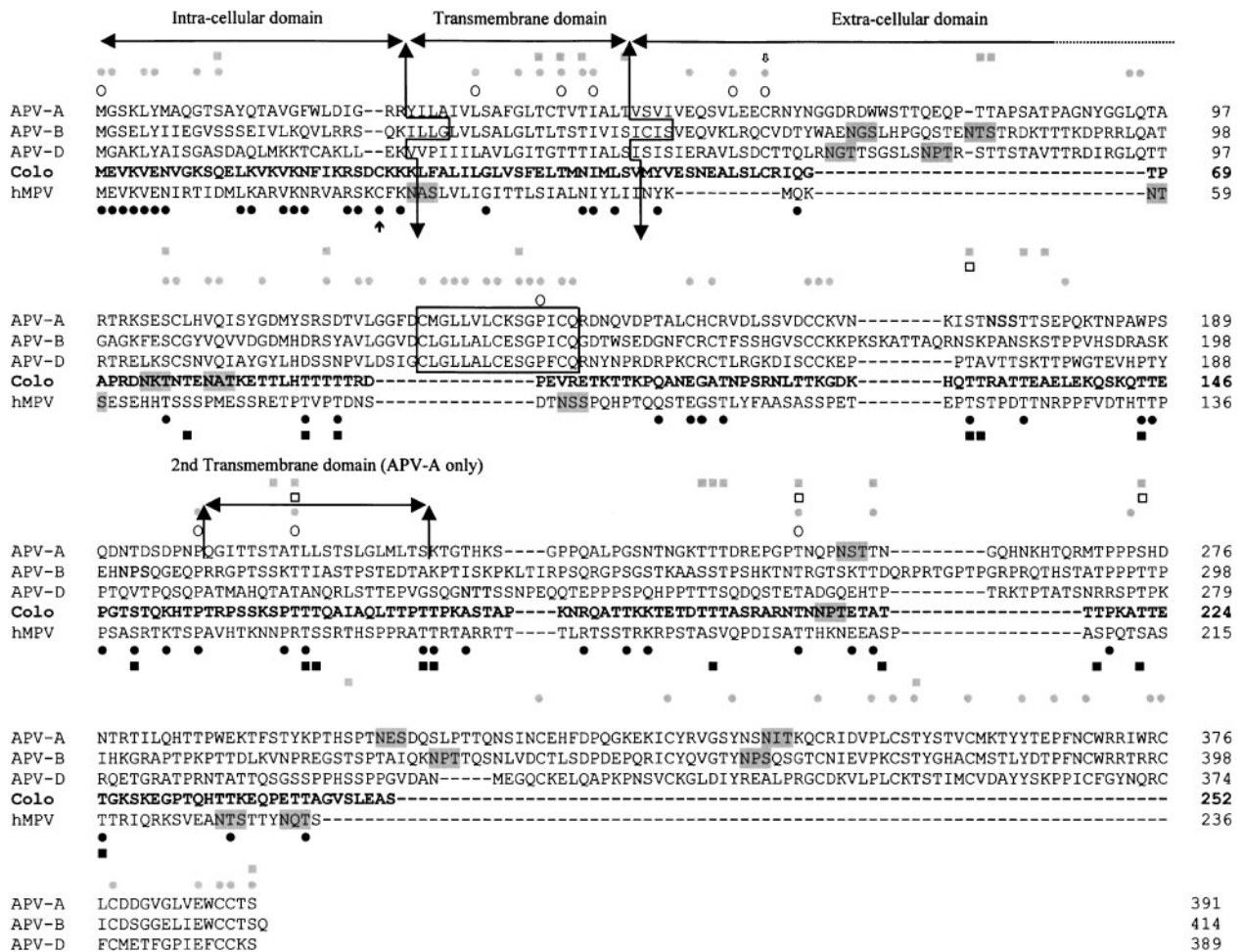


Fig. 2. Alignment of the predicted aa sequence of the G protein of the APV subgroup C Colorado isolate (Colo) with APV subgroup A strain CVL14/1, APV subgroup B strain 2119B, APV subgroup D strain Fr/85/1 and hMPV strain 00-1 (hMPV) (see accession numbers in the legend of Fig. 4). The alignment was done with CLUSTAL W and was presented so as to give the optimum alignment of Colorado with APV subgroups A, B and D on the one hand and of Colorado with hMPV on the other hand. Proposed intracellular, transmembrane and extracellular domains are indicated above the sequences. Open marks are used for comparison between APVs and Colorado, grey marks for comparison between APV subgroups A, B and D and closed marks for comparison between Colorado and hMPV. Circles indicate the positions of conserved aa residues. Squares indicate conserved serine and threonine residues (potential sites of O-glycosylation) and additional potential sites of O-glycosylation due to S/T exchanges and arrows indicate conserved cysteines. Predicted sites of N-glycosylation are enclosed in shaded boxes. The 15 aa stretch shown previously to be conserved between APV subgroups A, B and D (Báyon-Auboyer *et al.*, 2000) is boxed. Gaps (periods) introduced into the sequence to optimize alignments are not taken into account in the arbitrary numbering.

to ORF2b in APV subgroup B and to ORF3 in APV subgroup D (E values = 4×10^{-4} and 2×10^{-5} , respectively). These similarities corresponded to, at most, 33.8% aa identity in pairwise alignments (between the Colorado ORF2a and APV subgroup B ORF2b polypeptides), with, at most, three consecutive conserved aa. They were much less significant than found between the polypeptides encoded by ORF2a in APV subgroups A, B and D (maximum E value = 7×10^{-25} , at least 45.9% aa identity in pairwise alignments and the presence of stretches of at least 10 consecutive conserved aa). Searches based on the nt sequences of the

secondary ORFs of the Colorado virus revealed no positive matches with either hMPV or other RNA viruses.

Partial sequencing of the L gene

The gene start signal of the Colorado L gene was located only 3 nt after the polyadenylation signal of the putative G gene (versus a 209 nt long G–L intergenic region in hMPV; van den Hoogen *et al.*, 2002). The gene start of the L gene was GGACCAAGT, which differed from the typical APV gene start but also from the L gene starts determined previously in APV subgroups A (AGGACCAAT; Randhawa *et al.*, 1996)

and D (GGGACCAGT; Bâyon-Auboyer *et al.*, 2000) and finally from the hMPV L gene start (GAGACAAAT; van den Hoogen *et al.*, 2002). The start codon of the L gene was located 5 nt downstream of the gene start signal. The first 1104 nt located at the 5' extremity of the ORF was sequenced. This region exhibited 69.7% nt identity with the corresponding region in the hMPV L gene but only 59.4, 59.4 and 59.0% nt identity with the corresponding regions of the L genes in APV subgroup A (accession no. APU65312), HRSV (accession no. M75730) and bovine RSV (BRSV) (accession no. AF092942), respectively. The deduced protein of the Colorado isolate (Fig. 3) shared 73.6, 54.9, 40.8 and 39.4% aa identity with the L proteins of hMPV, APV subgroup A, HRSV and BRSV, respectively. Alignment of the N-terminal part of these polymerases revealed four highly conserved areas, corresponding to aa 9–32, 202–220, 270–299 and 353–368 in the Colorado L protein. These four areas (Fig. 3, grey boxes) proved to be 87.5, 63.2, 59.4 and 68.9% conserved in the *Pneumovirinae* tested and 95.8, 84.2, 90.6 and 87.5% conserved in the MPVs tested. The region spanning aa 353–368 corresponded to the first conserved domain ('domain I'), as defined by Stec *et al.* (1991), in the polymerases of negative-stranded RNA viruses.

Phylogenetic analysis

Phylogenetic analyses were performed with the G and L genes only (Fig. 4), as no significant alignment with the SH genes of other pneumoviruses could be produced.

Phylogenetic analysis of the L gene confirmed that the Colorado virus was more related to other MPVs (APV subgroup A and hMPV), as these three viruses clustered together in 85.6% of the bootstrap trees generated, than to members of the genus *Pneumovirus*. The Colorado virus and hMPV proved to be related in 72% of the bootstrap trees generated. Phylogenetic analysis of the G genes revealed two clusters with a bootstrap value higher than 75%: one grouped the RSVs from man and ungulates (79.4% bootstrap, with a significant subcluster due to the genetic relatedness of ORSV and BRSV), the other grouped all MPVs. Significant subclustering was apparent among MPVs, as isolates belonging to APV subgroups A, B and D clustered together and independently of both the Colorado isolate and hMPV in 95.8% of the bootstrap trees generated.

DISCUSSION

This study reports the first sequences of the putative SH and G genes (as defined by their position in the virus genome) and the partial sequence of the L gene of a subgroup C APV (Colorado isolate; Senne *et al.*, 1997). Whether the passage history of this isolate (which had been passaged repeatedly in CEFs) may contribute to some of the divergences in the nt sequences that are reported here is not known, and sequencing of primary APV subgroup C isolates is still required.

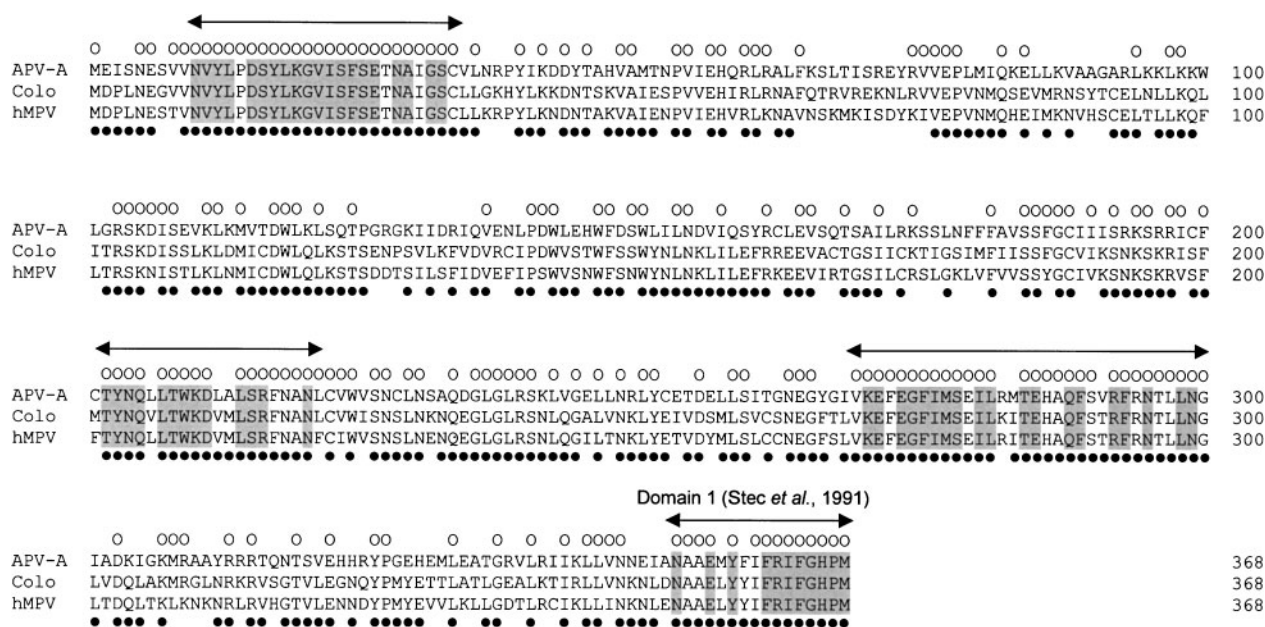


Fig. 3. Alignment of the predicted aa sequence of the NH₂-terminal extremity of the L polymerase protein of the APV subgroup C Colorado isolate (Colo) with APV subgroup A strain CVL/14.1 and hMPV strain 00-1 (hMPV) (see accession numbers in the legend of Fig. 4). The alignment was done and presented as in Fig. 1. Open circles are used for comparison between APV subgroup A and Colorado and closed circles are used for comparison between Colorado and hMPV. Circles indicate conserved residues. Double-headed arrows above the sequences identify long aa stretches with more than 84% conservation between MPVs (Colorado, hMPV and APV subgroup A), HRSV subgroup A and BRSV. Within these areas, shaded residues are conserved in all viruses.

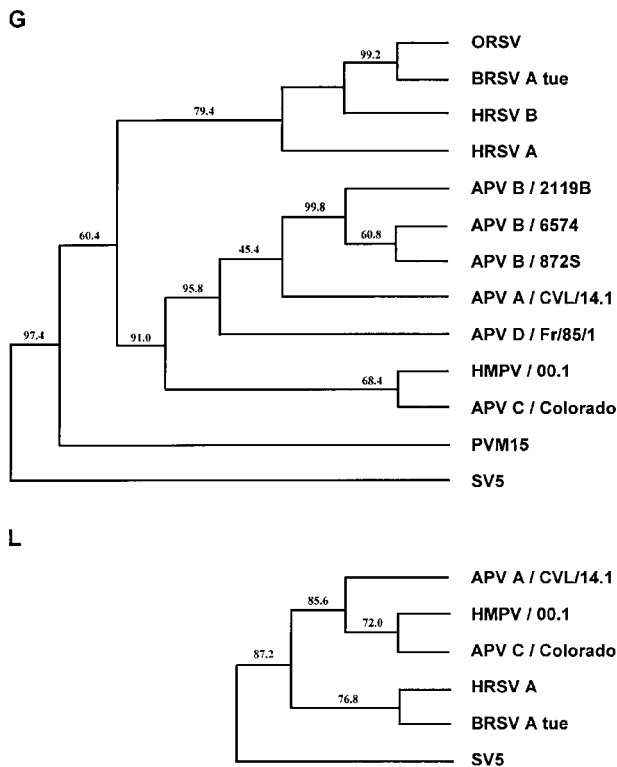


Fig. 4. Phylogenetic relationships between APV, hMPV and members of the genus *Pneumovirus*, based on the analysis of the nt sequence of the attachment glycoprotein gene (G) and the polymerase gene (L). Consensus phylogenetic trees were deduced from 500 trees generated by SEQBOOT. Each dataset was analysed using the neighbour-joining method. In each tree, the nt sequence of SV5 (accession no. AF052755) was used as an outgroup. Sites including gaps introduced by the outgroup were eliminated from the multiple alignment. Bootstrap percentages were based on the 500 bootstrap iterations. Branch length has no particular significance. When relevant, the virus subgroup and the reference of the strain have been indicated before the virus name. Accession numbers for the sequences included in the analysis are as follows: G tree, ORSV (L08470), BRSV subgroup A strain tue (AF092942), HRSV subgroup B (M17213), HRSV subgroup A (Z33423), APV subgroup B strain 2119b (L34031), APV subgroup B strain 6574 (L34033), APV subgroup B strain 872S (L34034), APV subgroup A strain CVL/14.1 (SA40185), APV subgroup D strain Fr/85/1 (AJ251085), hMPV strain 00-1 (AF371337) and PVM strain 15 (D11129). L tree: APV subgroup A strain CVL/14.1 (APU65312), hMPV strain 00-1 (AF 371337), HRSV subgroup A (AF254574) and BRSV (AF065167).

The conserved features in the Colorado and hMPV putative SH proteins included a similar hydrophobicity, conserved cysteine residues in the putative extracellular domain and short conserved stretches of aa. Another striking conserved feature in the SH gene was the location of a stop codon several codons upstream of the typical transcription termination sequence (24 untranslated codons including two additional in-frame stop codons in Colorado versus

17 untranslated codons with two additional in-frame stop codons in hMPV; van den Hoogen *et al.*, 2002). Such an organization is not found in the APV subgroup A SH gene, where no typical APV gene end signal (AGTTA) is apparent, although a poly(A) tract occurs 15 codons downstream of the stop codon (Ling *et al.*, 1992). One possibility might be that truncated SH proteins have been produced in APV subgroup C and hMPV by the introduction of premature stop codons in longer ancestor sequences. A similar mechanism has been described in the G gene of HRSV mutants that produce truncated forms of the G protein (Rueda *et al.*, 1991). Further studies are needed to determine if APV subgroups B and D, which are more genetically related to APV subgroup A than to APV subgroup C (Shin *et al.*, 2002; Seal, 1998; B  yon-Auboyer *et al.*, 2000), exhibit an APV subgroup C- or -A-like SH organization or, possibly, subgroup-specific SH changes. The biological significance of differences in the SH proteins is still largely unknown in the *Pneumovirinae*. Between HRSV subgroups A and B, the SH protein is mostly conserved (71.9% aa identity between representatives of the two subgroups; Chen *et al.*, 2000) and the short HRSV SH protein (64 and 65 aa in subgroups A and B, respectively) is suspected to have its integrity maintained due to an evolutionary or biological pressure (Chen *et al.*, 2000). Experiments with HRSV and BRSV mutants or recombinant viruses deleted in their SH gene (Karron *et al.*, 1997; Bukreyev *et al.*, 1997; Techaarpornkul *et al.*, 2001; Karger *et al.*, 2001) showed that SH is dispensable for virus growth *in vitro* but that a lack of SH may, in a murine experimental model, selectively impair HRSV replication in the upper respiratory tract and not in the lung (Bukreyev *et al.*, 1997). Whether a lower biological pressure in APV explains a greater variation among the SH genes of different APV subgroups or whether different SH proteins have a critical role in the host range, pathogenesis or organ tropism of MPVs awaits further investigation. The fact that extracellular cysteine residues are conserved in the Colorado, hMPV and APV subgroup A SH proteins, however, suggests that the secondary structure might be critical for their biological activity.

The putative attachment glycoprotein of APVs, G, is the most variable protein between APV subgroups (Juh  sz & Easton, 1994; B  yon-Auboyer *et al.*, 2000) and is believed to represent the basis for antigenic variation. The determination of the nt sequence of the G gene of an APV subgroup C virus was hence of major interest, as this subgroup has been shown to be the more antigenically divergent among APVs (Cook *et al.*, 1999; Toquin *et al.*, 1999b, 2000) and to escape molecular identification tests based on subgroup-specific oligonucleotide primers defined in the G gene (B  yon-Auboyer *et al.*, 1999). The G sequence reported here corroborates these previous findings, as the Colorado putative G gene and its deduced protein appeared to share very low nt and aa relatedness with other APV subgroups. Indeed, the best aa identity of the putative G protein of subgroup C was 21.0% with APV subgroup A (versus at least 28.1% between other APV subgroups;

Bäyon-Auboyer *et al.*, 2000). Consistent with this low aa identity, the Colorado putative G protein lacked several typical features recognized previously in its A, B and D counterparts, such as the highly conserved extracellular cysteine residues and the short hydrophobic and conserved aa stretch located in the G extracellular domain (Juhasz & Easton, 1994; Bäyon-Auboyer *et al.*, 2000).

Interestingly, the Colorado G protein also shared only 20·6 % overall aa identity with its hMPV counterpart, with an even lower conservation of the extracellular domain. This is less than what is found between RSV isolates from different mammalian species (25 to 29 % aa identity exists between G from HRSV subgroups A and B, BRSV subgroup A and ORSV). In spite of this low conservation, the Colorado putative G gene shared with its hMPV counterpart a highly conserved putative intracellular domain of its ORF1 product and the presence of shorter secondary ORFs. The phylogenetic analysis based on the G nt sequences produced results that are consistent with these observations: among MPVs (91·0 % bootstrap value), the Colorado and hMPV G genes appear both significantly different from the A, B and D APV genes (which are grouped in 95·8 % of the bootstrap trees generated) and only weakly related to the other (68·4 % bootstrap value). These results should, however, be confirmed once more hMPV and APV subgroup C-related sequences become available, as more sequences would ensure that the alignment of the aa sequences (first step in the phylogenetic analysis) is indeed relevant. Finally, a specific feature of the Colorado G protein, as compared with other MPVs, was the presence in its predicted transmembrane domain of a methionine codon followed by a predicted signal peptidase cleavage site. In HRSV, due to a similar organization, translation may initiate within the transmembrane domain, at the second ATG codon, leading to the production of a G protein with a truncated anchor. Following cleavage by a signal peptidase, Gs, a soluble form of G, is eventually secreted (Hendricks *et al.*, 1987; Roberts *et al.*, 1994; Lichtenstein *et al.*, 1996). Hence, the possibility that the Colorado isolate might also produce a soluble secretory form of G should also be considered.

Both the Colorado and hMPV G genes include potential secondary ORFs. In the Colorado isolate, most of these were not flanked by the typical gene start and gene end sequences and they did not include any start codon either. However, as observed by van den Hoogen *et al.* (2002), it cannot be ruled out that some of these ORFs are used in the coding of as yet unrecognized virus proteins due to some editing events. In this respect, the ORF3-encoded 137 aa peptide might be especially interesting, as it is the only Colorado secondary ORF to end with AGTTA, a typical MPV gene end signal (overlapping with its TAG stop codon, as found in several APV genes). In APVs belonging to subgroups A, B or D, there has been no experimental evidence so far that polypeptides different from the ORF1-encoded protein might be encoded by the G gene. However, the G genes of subgroup A, B and D viruses all contain a potential ORF2,

which is devoid of any gene start or gene end signals but exhibits a better inter-subgroup conservation than the ORF1-encoded protein (at least 45·9 % aa identity between the ORF2 polypeptide versus, at best, 35·1 % between the G proteins; Bäyon-Auboyer *et al.*, 2000).

Finally, the Colorado L gene was partially sequenced. As already observed with subgroup A APV (Randhawa *et al.*, 1996), subgroup D APV (Bäyon-Auboyer *et al.*, 2000) and hMPV (van den Hoogen *et al.*, 2002), the L gene start differed from those of the upstream genes. The fact that various changes in the L gene start do occur in different viruses might indicate that these changes contribute to confer a different level of expression to the L gene, as compared with other genes that include a more typical gene start. Highly conserved aa blocks were apparent in the N-terminal part of the polymerases of the *Pneumovirinae*, closer to the N-terminal end of the protein than domain I (Stec *et al.*, 1991). Such conserved domains correspond to those defined previously by N. Tordo and others as the 'NH₂-terminal' and 'pre-1' conserved domains (N. Tordo, Pasteur Institute, Paris, France, personal communication). The identification of such aa domains, downstream of the G gene end and conserved in all MPVs (and thus likely to be highly conserved at the nt level), is an important step towards the definition of conserved oligonucleotide primers that could be used to amplify the possibly highly divergent G gene of as yet unrecognized MPVs in other animal species or in other genetic lineages of hMPV (Peret *et al.*, 2002; van den Hoogen *et al.*, 2002).

Altogether, the present findings show that the Colorado isolate (proposed as the type species for subgroup C APV; Seal, 2000) and the hMPV isolated recently (van den Hoogen *et al.*, 2001, 2002) share in their SH and G genes some features that make them clearly different from the APV subgroups known previously. However, the two viruses also exhibit extensive sequence divergence from one another (26·9 and 21 % aa identity in their SH and G genes, respectively, versus 68–88 % aa identity in their N, P, M, F and M2 genes). Several genetic lineages of hMPV are presently under study (Peret *et al.*, 2002; van den Hoogen *et al.*, 2002) and it cannot be ruled out that some as yet unrecognized hMPV genetic lineage might prove more genetically related to APV subgroup C than hMPV isolate 00-1. However, the level of divergence reported here makes it unlikely that the common ancestor of APV subgroup C and hMPV viruses crossed the species barrier in recent history. Regarding the classification of MPVs, the unique antigenicity of the Colorado virus, together with the different organization of its G gene, suggest that this virus might represent a second APV serotype among the MPVs. Based on the extensive divergence in their putative G proteins, antigenic differences between hMPV and APV subgroup C might be expected. Whether hMPV represents a third serotype within the genus *Metapneumovirus* or whether it represents a different subgroup within a possibly 'APV subgroup C defined' second MPV serotype, awaits an

extensive evaluation of the cross-reactivity of these viruses in antigenic studies.

NOTE ADDED IN PROOF

A recent paper (Alvarez *et al.*, *J Clin Microbiol* **41**, 1730–1735) reports on the reverse transcription, amplification and sequencing of a 1321 nt long G-encoding mRNA derived from cells infected by the Colorado virus or by several APV subgroup C isolates from the United States. The sequences of the 5' and 3' ends of this mRNA are identical to the sequences reported here from genomic RNA for the SH and G genes, respectively. Such a long mRNA could be transcribed if APV subgroup C polymerase performs some readthrough at the SH–G gene junction.

ACKNOWLEDGEMENTS

The authors wish to thank Dr Andrew Easton (University of Warwick, UK) for helpful discussions and kind review of the manuscript, and Dr Noël Tordo (Pasteur Institute, Paris, France) for his help with the conserved polymerase domains.

REFERENCES

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
- Bäyon-Auboyer, M. H., Jestin, V., Toquin, D., Cherbonnel, M. & Eterradossi, N. (1999). Comparison of F-, G- and N-based RT-PCR protocols with conventional virological procedures for the detection and typing of turkey rhinotracheitis virus. *Arch Virol* **144**, 1091–1109.
- Bäyon-Auboyer, M. H., Arnault, C., Toquin, D. & Eterradossi, N. (2000). Nucleotide sequences of the F, L and G protein genes of two non-A/non-B avian pneumoviruses (APV) reveal a novel APV subgroup. *J Gen Virol* **81**, 2723–2733.
- Bell, I. G. & Alexander, D. J. (1990). Failure to detect antibody to turkey rhinotracheitis virus in Australian poultry flocks. *Aust Vet J* **67**, 232–233.
- Bukreyev, A., Whitehead, S. S., Murphy, B. R. & Collins, P. L. (1997). Recombinant respiratory syncytial virus from which the entire SH gene has been deleted grows efficiently in cell culture and exhibits site-specific attenuation in the respiratory tract of the mouse. *J Virol* **71**, 8973–8982.
- Buys, S. B. & Du Preez, J. H. (1980). A preliminary report on the isolation of a virus causing sinusitis in turkey in South Africa and attempts to attenuate the virus. *Turkeys* **28**, 36–46.
- Chen, M. D., Vasquez, M., Kahn, J. S. & Buonocore, L. (2000). Conservation of the respiratory syncytial virus SH gene. *J Infect Dis* **182**, 1228–1233.
- Claros, M. G. & von Heijne, G. (1994). TopPredII: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10**, 685–686.
- Collins, M. S., Gough, R. E. & Alexander, D. J. (1993). Antigenic differentiation of avian pneumovirus isolates using polyclonal antisera and mouse monoclonal antibodies. *Avian Pathol* **22**, 469–479.
- Cook, J. K. (2000). Avian rhinotracheitis. *Rev Sci Tech* **19**, 602–613.
- Cook, J. K. A., Jones, B. V., Ellis, M. M., Li, J. & Cavanagh, D. (1993). Antigenic differentiation of strains of turkey rhinotracheitis virus using monoclonal antibodies. *Avian Pathol* **22**, 257–273.
- Cook, J. K. A., Huggins, M. B., Woods, M. A., Orbell, S. J. & Mockett, A. P. A. (1995). Protection provided by a commercially available vaccine against different strains of turkey rhinotracheitis virus. *Vet Rec* **136**, 392–393.
- Cook, J. K. A., Huggins, M. B., Orbell, S. J. & Senne, D. A. (1999). Preliminary antigenic characterization of an avian pneumovirus isolated from commercial turkeys in Colorado, USA. *Avian Pathol* **28**, 607–617.
- Eterradossi, N., Toquin, D., Guittet, M. & Bennejean, G. (1995). Evaluation of different turkey rhinotracheitis viruses used as antigens for serological testing following live vaccination and challenge. *Zentralbl Veterinarmed B* **342**, 175–186.
- Felsenstein, J. (1993). PHYLIP: Phylogenetic Inference Package, version 3.52c. Department of Genetics, University of Washington, Seattle, WA, USA.
- Freytmuth, F., Eterradossi, N., Toquin, D., Jestin, V., Vabret, A., Petitjean, J. & Gouarin, S. (2002). First detection of human metapneumovirus in hospitalized children in France. *Virologie* **6**, S14 (in French).
- Giraud, P., Bennejean, G., Guittet, M. & Toquin, D. (1986). Turkey rhinotracheitis in France: preliminary investigations on a ciliostatic virus. *Vet Rec* **119**, 606–607.
- Gupta, R., Jung, E. & Brunak, S. (2002). Prediction of N-glycosylation sites in human proteins (<http://genome.cbs.dtu.dk/services/NetNGlyc/>)
- Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. & Brunak, S. (1998). NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj J* **15**, 115–130.
- Hendricks, D. A., Baradaran, K., McIntosh, K. & Patterson, J. L. (1987). Appearance of a soluble form of the G protein of respiratory syncytial virus in fluids of infected cells. *J Gen Virol* **68**, 1705–1714.
- Juhász, K. & Easton, A. J. (1994). Extensive sequence variation in the attachment (G) protein gene of avian pneumovirus: evidence for two distinct subgroups. *J Gen Virol* **75**, 2873–2880.
- Karger, A., Schmidt, U. & Buchholz, U. J. (2001). Recombinant bovine respiratory syncytial virus with deletions of the G or SH genes: G and F proteins bind heparin. *J Gen Virol* **82**, 631–640.
- Karron, R. A., Buonagurio, D. A., Georgiu, A. F. & 8 other authors (1997). Respiratory syncytial virus (RSV) SH and G proteins are not essential for viral replication *in vitro*: clinical evaluation and molecular characterization of a cold-passaged, attenuated RSV subgroup B mutant. *Proc Natl Acad Sci U S A* **94**, 13961–13966.
- Lichtenstein, D. L., Roberts, S. R., Wertz, G. W. & Ball, L. A. (1996). Definition and functional analysis of the signal anchor domain of the human respiratory syncytial virus glycoprotein G. *J Gen Virol* **77**, 109–118.
- Ling, R., Easton, A. J. & Pringle, C. R. (1992). Sequence analysis of the 22K, SH and G genes of turkey rhinotracheitis virus and their intergenic regions reveals a gene order different from that of other pneumoviruses. *J Gen Virol* **73**, 1709–1715.
- McDougall, J. S. & Cook, J. K. A. (1986). Turkey rhinotracheitis: preliminary investigations. *Vet Rec* **118**, 206–207.
- Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1–6.
- Nissen, M. D., Sieberg, D. J., Mackay, I. M., Sloots, TP. & Withers, S. J. (2002). Evidence of human metapneumovirus in Australian children. *Med J Aust* **176**, 188.
- Peret, T. C., Boivin, G., Li, Y., Couillard, M., Humphrey, C., Osterhaus, A. D., Erdman, D. D. & Anderson, L. J. (2002).

Characterization of human metapneumoviruses isolated from patients in North America. *J Infect Dis* **185**, 1660–1663.

Pringle, C. R. (1998). Virus taxonomy – San Diego 1998. *Arch Virol* **143**, 1449–1459.

Randhawa, J. S., Wilson, S. D., Tolley, K. P., Cavanagh, D., Pringle, C. R. & Easton, A. J. (1996). Nucleotide sequence of the gene encoding the viral polymerase of avian pneumovirus. *J Gen Virol* **77**, 3047–3051.

Roberts, S. R., Lichtenstein, D. L., Ball, L. A. & Wertz, G. W. (1994). The membrane-associated and secreted forms of the respiratory syncytial virus attachment glycoprotein G are synthesized from alternative initiation codons. *J Virol* **68**, 4538–4546.

Rueda, P., Delgado, T., Portela, A., Melero, J. A. & Garcia-Barreno, B. (1991). Premature stop codons in the G glycoprotein of human respiratory syncytial viruses resistant to neutralization by monoclonal antibodies. *J Virol* **65**, 3374–3378.

Seal, B. S. (1998). Matrix protein gene nucleotide and predicted amino acid sequence demonstrate that the first US avian pneumovirus isolate is distinct from European strains. *Virus Res* **58**, 45–52.

Seal, B. S. (2000). Avian pneumoviruses and emergence of a new type in the United States of America. *Anim Health Res Rev* **1**, 67–72.

Senne, D. A., Edson, R. K., Pedersen, J. C. & Panigraphy, B. (1997). Avian pneumovirus update. In *Proceedings of the 134th Annual Congress of the American Veterinary Medical Association*, p. 190. Reno, NV, USA.

Shin, H. J., Cameron, K. T., Jacobs, J. A. & 8 other authors (2002). Molecular epidemiology of subgroup C avian pneumoviruses isolated in the United States and comparison with subgroup A and B viruses. *J Clin Microbiol* **40**, 1687–1693.

Stec, D. S., Hill, M. G., III & Collins, P. L. (1991). Sequence analysis of the polymerase L gene of human respiratory syncytial virus and predicted phylogeny of nonsegmented negative-strand viruses. *Virology* **183**, 273–287.

Techaarpornkul, S., Barretto, N. & Peebles, M. E. (2001). Functional analysis of recombinant respiratory syncytial virus deletion mutants

lacking the small hydrophobic and/or attachment glycoprotein gene. *J Virol* **75**, 6825–6834.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Toquin, D., Eterradossi, N., Guittet, M. & Bennejean, G. (1992). Infectious rhinotracheitis in turkeys: antigenic differences revealed by ELISA. In *European Commission Proceedings: New and Evolving Pathogens Virus Diseases of Poultry*, pp. 111–121. Edited by M. S. McNulty & J. B. McFerran. Publication VI/4104/94-EN, III-121.

Toquin, D., Eterradossi, N. & Guittet, M. (1996). Use of a related ELISA antigen for efficient TRT serological testing following live vaccination. *Vet Rec* **139**, 71–72.

Toquin, D., Băyon-Auboyer, M. H., Jestin, V., Eterradossi, N. & Morin, H. (1999a). Isolation of a pneumovirus from a Muscovy duck. *Vet Rec* **145**, 680.

Toquin, D., Băyon-Auboyer, M. H., Jestin, V. & Eterradossi, N. (1999b). Serological response and cross protection following infection by a non-A, non-B strain of turkey infectious rhinotracheitis virus. *Proceedings Troisième Journées de la Recherche Avicole*, pp. 221–224 (St Malo, France, 23–25 March 1999) (in French).

Toquin, D., Băyon-Auboyer, M. H., Senne, D. A. & Eterradossi, N. (2000). Lack of antigenic relationship between French and recent North American non-A/non-B turkey rhinotracheitis viruses. *Avian Dis* **44**, 977–982.

van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A. & Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med* **7**, 719–724.

van den Hoogen, B. G., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. (2002). Analysis of the genomic sequence of a human metapneumovirus. *Virology* **295**, 119–132.

Yu, Q., Davis, P. J., Li, J. & Cavanagh, D. (1992). Cloning and sequencing of the matrix protein (M) gene of turkey rhinotracheitis virus reveal a gene order different from that of respiratory syncytial virus. *Virology* **186**, 426–434.